DOCUMENT RESUME

ED 058 315 TM 001 030

AUTHOR Skager, Rodney
TITLE The System for Objectives-Based Evaluation--Reading.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Cooperative Research Program.
PUB DATE Sep 71
NOTE 6p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Curriculum Planning; *Educational Accountability; Educational Improvement; Educational Objectives; Educational Programs; *Evaluation Methods; *Information Systems; Instructional Programs; *Program Evaluation; Reading Improvement; *Reading Programs; Standardized Tests; Testing Programs; Test Interpretation
IDENTIFIERS System for Objectives-Based Evaluation--Reading

ABSTRACT

The System for Objectives-Based Evaluation--Reading (SOBE-R) is discussed. Evaluative inadequacies of standardized tests are listed, and the functions and components of the SOBE-R program are examined. Preliminary evaluations of SOBE-R are included, as are its expected accomplishments. (DLG)

ED 058315

THE SYSTEM FOR OBJECTIVES-BASED EVALUATION — READING

Rodney Skager

University of California, Los Angeles

The Need for SOBE-R

The System for Objectives-Based Evaluation — Reading (SOBE-R) is a child of the movement toward rigorous evaluation and accountability in American education. As such, it is intended as a tool for evaluating reading programs and managing instruction in the classroom. In a more fundamental sense, the development of SOBE-R will provide similar information systems in other areas of instructional content.

Standardized tests such as the *Cooperative Reading Test*, *Stanford Achievement Test*, etc., now widely used in evaluation, are virtually useless for program evaluation and educational accountability, just as they have been similarly ineffective as tools for the management of instruction in the classroom. Documentation of these statements is available in the literature, in earlier reports from CSE, and in the frequently expressed views of teachers, administrators, and professionals involved in evaluative activities at all levels.

Standardized testing programs are inadequate for the evaluation of instructional programs and the determination of educational accountability for several reasons:

(1) *Standardized tests ordinarily have a low degree of overlap with the actual objectives of instruction at any given time or place.* Standardized tests, except for some individually administered diagnostic instruments, are designed to measure highly generalized skills. This inevitably means that scores on such tests have only a distant, even tenuous, relationship to the outcomes of the real-life instruction in a given situation. One cannot determine the effectiveness of educational programs or establish accountability in any meaningful sense if it is not clear that the devices used to assess educational outcomes actually measure the objectives a given teacher, school, district, or state is attempting to accomplish over a defined period of time.

TM 001 030

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

(2) *Standardized tests are not useful as aids in the planning of instruction, particularly the specification of instructional objectives.* One would hardly examine the *Cooperative Reading Test*, used in the California state mandated testing program, in order to identify specific skills which might become objectives guiding curriculum development. Because the test must be administered in a single sitting, only a few complex or "terminal" objectives can be represented in test content, particularly those of a type that are conveniently assessed in a brief group administration.

(3) *Standardized tests often require skills or aptitudes that may be influenced to only a limited degree by experiences in the classroom.* Standardized group tests frequently incorporate items or tasks highly similar to those found in measures of aptitude and intelligence. The capacities such materials reflect are rarely if ever the direct object of classroom instruction. This means that such tests are literally programmed to be insensitive to instruction, particularly over the short range, and are potentially misleading in determining the effectiveness of teaching.

(4) *Standardized tests do not indicate the extent to which individuals or groups of students have mastered the spectrum of instructional objectives.* Designed to spread students out on a continuum with respect to the skills measured by the test, standardized tests lead to normative rather than mastery or criterion-referenced interpretations. Normative interpretations are entirely relative to the individuals who are selected for the norming group, rather than to pre-specified standards or goals which schools might be committed to accomplishing at various levels of the educational process. Using such dull tools for "documenting" accountability constitutes a hazard for educational institutions and the people who work in them, particularly in the context of contemporary public misunderstanding of the meaning of test norms. The fact that students in a school score above or below average with respect to some ill-defined normative group tells nothing whatsoever about what the school has accomplished with respect to its objectives.

(5) *The use of standardized achievement tests in the socio-political context of "evaluation" and "accountability" has and will lead to corruption and dishonesty among educational professionals and to the further erosion of public trust in the schools and the people who run them.* This is a serious issue. It is already the subject of at least one legal action in the nation's second largest school system. Faced with public pressure that is often in itself irrational and destructive, it is all too easy for subtle pressure to be applied to principals and teachers to "prepare students for the assessment" by teaching students responses to specific test items rather than by developing the underlying skills which these items reflect. Standardized tests are readily available at all levels of any school system, and are brief enough to be highly susceptible to coaching. Test security and control may be feasible for programs like the College Boards or the American College Testing Program, where representatives of the testing agency handle the assessment and the examinees come to a central location. Similar controls are not feasible in large scale evaluations for school programs.

An evaluation system which can avoid the security problem and successfully deal with the other criticisms listed above is one which (a) permits the assessment of a wide variety of specific learning objectives selected because they are isomorphic with the program under scrutiny by means of (b) a test or item-sampling strategy which assures that the particular items given to a student are not known in advance by those who will administer the test. SOBE-R is being developed to deal with the kinds of problems outlined above. Before describing the system itself, however, it might be helpful to summarize its intended functions.

The Functions of SOBE-R

Three, and possibly four basic roles are envisioned for SOBE-R.

The first two are already incorporated in field testing/development activities currently underway. The third is properly slated for the second phase of field testing, and the fourth is being studied in terms of its feasibility.

Curriculum Selection: We accept Gagné's view that curriculum is defined by precise specification of the behavioral outcomes of instruction. Unfortunately, actually writing such performance outcomes in concrete detail requires resources that are beyond the capacity of the vast majority of educational organizations. Schools lack personnel trained in developing performance objectives and assessment procedures congruent with these objectives. Funds for training or hiring qualified personnel are hard to come by. As a result, in many parts of the country classroom teachers have been pressed into this kind of activity, often in addition to their regular responsibilities, and usually with highly unsatisfactory results in terms of quality of materials and teacher morale.

The SOBE-R system is shaped by the idea that costly *development* of performance objectives on a local basis can be circumvented by making it possible to *select* objectives from a comprehensive and readily utilized file of objectives. The main emphasis in the field/testing/developmental activities of the project during the spring of 1971 has been directed at finding the most efficient and user-appropriate procedures for selecting SOBE-R objectives for an individual school, a school district, and a state.

Program Evaluation: Once instructional objectives have been selected for a school or larger administrative entity, an assessment system must be created for periodically determining the extent to which these objectives are reflected in student performance. We are referring here to assessments which provide information for planning and resource allocation at levels higher than the individual classroom, as well as reports to the public or their representatives as to the effectiveness of school programs. The present standardized testing programs of hundreds of school districts or the annual statewide testing in California are examples of ineffectual attempts to meet this need. Looking to the future the Florida State Department of Education has determined that two types of public reports will be developed from the information generated by the use of SOBE-R assessment materials and procedures in the coming year:

1. General summary reports ... "To inform the legislature, the State Board of Education, and the general public relative to the operations and findings of the Florida Assessment Program, and to make recommen-

Evaluation Comment – Page 7

dations for improvement."[1]

2. Reports to educators in the Department and in school districts identifying specific areas of achievement where reading programs have been successful or unsuccessful.

The report to the lay public and their representatives in the legislature should, in part, enable the reader to—

"...describe how his county compares with other counties and with the state as a whole in each goal area..."

"...state, in general terms, what kinds of things his district should concentrate on to improve instruction in reading..."

"...ask a classroom teacher meaningful questions about the way he teaches the skill described."

The above goals were abstracted from a memorandum circulated recently in the Division of Elementary and Secondary Education and are admittedly ambitious in their intent, but they do reflect the need perceived in Florida for information bearing on program evaluation.

The SOBE-R system is designed to meet this need by making available assessment materials congruent with each objective selected, as well as providing procedures for test administration, sampling of students, and summary of results.

Classroom Feedback: Periodic program evaluation does not satisfy all the potential information needs of the classroom from the point of view of the teacher. Program evaluation is used for decisions about instructional programs affecting students in relatively large aggregate groups. Teachers need means of assessing progress on a short term basis — say on one or at most a few performance objectives. Materials facilitating this type of classroom feedback can be derived from the SOBE-R objectives and test items. The trial development of procedures for implementing this aspect of SOBE-R will begin during Fall, 1971, as a part of the project's activities at an elementary school in Denver, Colorado.

Instructional Design: A fourth area under development in SOBE-R is entirely tentative at this point. It is based on the spontaneous requests of many of the teachers who selected objectives from the SOBE-R file during field testing in Spring, 1971. These individuals have expressed great interest in an additional component of the system which would index instructional materials and procedures to each objective. Such an additional component is a logical step, and one which could well significantly increase the impact of the system in the classroom. Further exploration of the feasibility and usefulness of this development is required.

The Components of SOBE-R

To understand the functions served by present and future field testing arrangements for SOBE-R it is necessary to bear in mind that the final system will incorporate at least ten components, some of which do not yet exist in concrete form. The latter are all in the software or utilization category, however.

It is also important to bear in mind that we view "field testing" at this stage of the development of SOBE-R in a somewhat different light than the term is ordinarily understood. In most instances we are actually involved in a process of "field development testing," especially with regard to utilization materials. In this sense our evaluative activities are basically formative in nature, involving planning, implementation, and progress evaluation. The specifics of some of these arrangements will be provided in detail below. One example occurred early in 1971 as a part of our relationship with the elementary school in Denver. In designing the Guide for Selecting Objectives (item 4 below) the project staff first met with a team of teachers and the principal and jointly discussed various strategies. The same teachers then utilized the procedures decided on to select objectives for their own school, reporting on the process periodically in writing as well as in later interviews. These experiences constituted the initial basis for writing a Guide for users which was then used on a larger scale in two school districts in California.

It cannot be emphasized too strongly that this form of developmental activity is evaluative in its fundamental nature, not only according to the model utilized by the Center for the Study of Evaluation, but in every other sensible conception of evaluation advanced in the last fifteen years. Therefore, the question addressed in this report is not primarily how SOBE-R *will* be evaluated, but rather how SOBE-R *is* being evaluated.

The following, then, are the parts that make up SOBE-R. The formative evaluation strategy appropriate to each is indicated as well.

(1) *Classification System:* Work on SOBE-R began with a content-analysis of the domain of reading leading to the development of a system for classifying all of the component reading skills. This task had to be accomplished both to ensure that the performance objectives would represent a complete set, and to provide a means of organizing for the eventual user the task of selecting objectives.

Although the system was useful in generating objectives, it has not proved to be without flaws as an aid to teachers selecting objectives from the SOBE-R file. We learned from on-site observations as well as reports by teachers that the extensive branching in the initial classification system was difficult to follow, and as a result some teachers simply reviewed the entire set of objectives without really using the classification system. Moreover, the classification system did not always facilitate finding particular types of objectives, as in the case, for example, where a selector was looking specifically for encoding objectives, or phonics objectives, or objectives relating to non-standard English. With these problems in mind, a revised classification system has been developed which (a) provides more basic categories and hence less branching from category heading to performance objective and (b) facilitates the identification of particular types of objectives of special interest. Revisions in the system have, in turn, suggested the development of some additional performance objectives, although the great majority of the existing objectives are simply being reclassified.

(2) *Performance Objectives:* The SOBE-R file now contains approximately 1000 performance objectives. Inevitably, most of the work on selection and revision of these objectives has been done by SOBE-R staff. A variety of specific suggestions as to revision or development of additional objectives has been generated by teachers in field-testing agencies as well as reading specialists in teacher-training institutions; usually these suggestions have been valuable. In addition, a questionnaire has been used to obtain the judgments of several hundred teachers and reading specialists in the two California school districts as to the completeness and quality of the set of objectives.

[1] *Plan for Educational Assessment in Florida.* Florida Department of Education, Tallahassee, Florida, Final Report, March, 1971.

One SOBE-R staff member has compared SOBE-R objectives with the goals of a number of curriculum guides in the reading area, looking for possible gaps or inconsistencies in our materials. The Florida Department of Education has recently provided us with the state curriculum guide in reading as well as the guides for two Florida school districts judged by the Department to be the most complete and comprehensive in the state. The results of this check on SOBE-R will be summarized in a separate report.

(3) *Assessment Materials:* Pools of assessment items are being constructed to measure each objective in the SOBE-R system. SOBE-R objectives have been deliberately written so as to allow some variation in item type, response format, etc., so that the pool of items under each objective does not amount to a repetition, with minor variations, of the same task. For example, wherever possible, we are developing for many objectives both selection and construction item formats (e.g., "selection' items require the respondent to *find* the correct response from a set of alternatives, while "construction" items require the production of a response with clues provided by alternatives).

Quality control in the production of so many assessment items has several facets. First, there must be formalized review procedures on individual items. Our own in-house review procedures require that all items be examined by individuals other than the original item writer.

No matter how carefully items are written and reviewed, there is still the need for empirical try-out. Empirical data aids in the search for misleading items, such as items which can be answered correctly without the requisite knowledge or skill the item was intended to measure, items which have one or more misleading alternatives, items which do not relate to other items presumably measuring the sample objective, etc. The field testing arrangements with the Denver school, the two school districts in California, the Children's TV Workshop, and the State of Florida are all designed in part to provide this type of data, primarily during the coming year.

Finally, a fundamental question arises when one is developing assessment materials to measure highly explicit instructional objectives. If the assessment is to be useful, then we must establish that the assessment items actually measure the skill defined by the objective, that is, that the items are *congruent* with the objectives. (In standardized testing the congruency problem has its analogue in "content-validity," ordinarily established in a subjective, judgmental fashion.) Because the determination of congruency is so important to the evaluation of SOBE-R materials, this problem has been the subject of a research study by one of our project staff, T.A. Dahl. His work involves using SOBE-R items and objectives in investigating both analytical and judgmental strategies for determining congruency. The results of this research project will be available in winter, 1971. A preliminary discussion of the project has recently been published by CSE.[1]

(4) *Guide for Selection of Objectives:* Users need to be informed about the purpose of SOBE-R. The incorporation of approximately 1000 performance objectives in the SOBE-R file also makes it important that procedures be developed which will enable users to select objectives in an efficient and systematic manner. Such a guide should make it easy to find objectives by providing an easily understood categorization of reading skills, plus a series of steps to be followed in selection. Procedures and materials for recording and summarizing the results are also required.

The guide for selection of objectives, with various revisions along the way, was developed after initial work at the Denver elementary school and used in 46 schools in the two California school districts. Briefly, the Guide recommends that the selection be conducted by a team from each participating school, with one teacher representing each grade (or other) level of instruction. The group proceeds through a two-stage process. First, the categorization system in (1) above is used to make an initial selection among goals immediately above the level of performance objectives (e.g., incorporating two or more performance objectives). Next, examining only those performance objectives classified under the goals that have been chosen, the group makes its final selection. This procedure was designed to preclude examining approximately 1000 performance objectives by dealing, in the initial step, with only about 90 goals at a higher level of abstraction. Thus, *only* those objectives classified under each selected goal category need be examined in the final stage. In both steps, however, it has been found useful to classify goals or objectives as either "minimal" (to be achieved by 90 to 95% of the students) at a given level of the instructional program, or as supplementary/enrichment at the same level. This approach appears to help teachers establish priorities within the set of objectives.

The selection guidelines are now under review based on information obtained from the questionnaire, individual interviews conducted in one of the California school districts, and questions and comments generated in meetings with personnel in both districts.

Components (5) through (10), described below, are in early stages of development. These components will be defined, but cannot be described in as much detail as components already in field-test stages.

(5) *Sampling strategies for various levels of aggregation:* SOBE-R is adaptable to strategies involving the sampling of both students and items (or tests), and sampling procedures are presently being developed primarily in cooperation with the Florida State Department of Education. It has already been pointed out that such sampling has the advantage of rendering coaching for particular items or tests profitless. Sampling also has the advantage of efficiency, meaning that where assessments are being made of aggregates — schools of average size and above, school districts, and states — all students need not be tested and the tests themselves can be quite brief. Such sampling strategies are, of course, especially appropriate for program evaluation; for example, assessments of instructional programs administered to aggregates of students larger than single classrooms.

(6) *Procedures for Assembling Tests:* Once the selection of objectives has been accomplished, procedures must be available describing how tests are to be assembled from SOBE-R item pools under parameters characterizing the various situations in which tests might be used. Whether or not students, items, or both are to be sampled is one such parameter. The number of objectives selected is another, as is the number of times tests would be administered in a school year. As data accumulate on SOBE-R items such information could also be built into the test assembly procedures. The development of this component has already

[1] CSE Working Paper No. 15. T.A. Dahl. *Toward an Evaluative Methodology for Criterion-Referenced Measures: Objective-Item Congruence*, May, 1971.

begun.

(7) *Procedures for Administration and Distribution:* All assessment programs require plans for administering and distributing tests. The level of complexity of such plans is, of course, dependent on the number of students tested, the number of institutions involved, and whether or not sampling strategies are used. SOBE-R materials relating to this issue will probably be differentiated for the single school, the school district, and the state, with separate contingencies for sampling versus non-sampling strategies within each. Such materials are at best general guidelines and could not be expected to anticipate all of the specifics of a given assessment.

(8) *Reporting Formats:* Guidelines for reporting assessment results must take into account the same types of contingencies described in (7) above. In addition, the audience receiving the report adds another important consideration. It should be noted that two types of reports are planned, one to the general public and their representatives, and one (or perhaps two) of a more technical nature for planners, administrators, and research personnel.

(9) *Guidelines for Use of Assessment Results at Aggregate Levels:* Given that assessment results are available at the school, district, and state level, how might the information be used in effecting necessary changes in instructional programs? That is, the availability of extensive and programmatic evaluation data does not mean that the information will be used in actual decisions about needs assessment and planning. In our view, field-testing arrangements are a major device for entering this admittedly complex area. While actual development of related materials and documentation of the results are slated for the final phases of the development of SOBE-R, a preliminary planning document is being prepared in cooperation with Florida state personnel. This document will probably be in the form of a scenario, describing or anticipating hypothetical conditions which will only be encountered after assessment results are available. Similar activities will be undertaken at other levels of aggregation, especially schools and school districts.

(10) *Guidelines for the Use of SOBE-R Materials and Results in the Classroom:* Krathwohl has distinguished between three types of instructional objectives.[3] At the first level are broad and abstract goals used primarily to assign priorities and communicate intentions. The educational goals contained in the CSE Elementary School Evaluation KIT and the broad categories of the SOBE-R classification system are of this type. Objectives at the second level represent the detailed specification of the many outcomes encompassed by the broad goals at the first level. Krathwohl referred to this second class of objectives as "...building blocks for curriculum and instrumentation." The approximately 1000 objectives in the SOBE-R collection fit this characterization. Finally, there are extremely specific objectives used to "...create instructional materials which are the operational embodiment of *one particular* (italics ours) route..." to teaching. In other words, the third type of objective is properly limited in applicability to a single instructional program or package. Such objectives define skills, perceptions, etc., which are intended outcomes of very small steps in a particular learning process. Assessment materials congruent with this level of objective have been termed "curriculum embedded tests" in instructional programs such as IPI. It would be inappropriate to use the tests from the IPI reading program in determining the effectiveness of, say, the SWRL or Sullivan reading instructional systems. This kind of evaluation is the province of a system like SOBE-R which assesses outcome objectives not tied to any particular approach to instruction.

Krathwohl's distinctions between levels of objectives are useful in placing in perspective the ways in which the SOBE-R system will be useful to the classroom teacher as well as the ways in which it will not. First, teachers should be familiar with the objectives selected by the school, district, or other organizational entity of which they are a part, especially those objectives that define the curriculum at the particular grade (or other) level where they are responsible for instruction. In addition, they should be familiar with the nature of the measurement devices used to assess each objective. SOBE-R materials can and should be used for this purpose — that is, to define for the individual teacher the expectations of the larger system, the "Accountable for what?" question.

On the other hand, it is not likely that the results of assessment of a large school, district, or state utilizing sampling techniques will be of great value to the individual teacher, whose particular classroom may not even have been sampled or, even if it has, is not isolated in the data. This is certainly the case for the type of assessment system being developed in the state of Florida. The information SOBE-R will generate in this case relates primarily to allocation of resources and the planning of curricula for aggregates larger than the single classroom. In a different approach, SOBE-R assessment instruments, administered at the beginning and end of the school year to all children, as is planned in both of the California school districts working with us, can provide the teacher with feedback about entry skills and the progress made by individual students over the year.

Finally, whether the data-base of the assessment is a sample or a population, teachers should be able to use the objectives to help in planning their own instructional sequences, selecting materials, and developing their own curriculum embedded tests. In this regard, Florida State Department of Education earlier requested that CSE participate in a joint effort to develop training materials that will help teachers utilize the objectives in planning and teaching. The present commitments of the staff on the SOBE-R project involve other priorities during the coming year. However, other agencies, particularly the New York State Department of Education, have already begun to attack the problem of teacher utilization of objectives and assessment results. Exchange of materials and cooperation in field-testing activities between NYSDE and the Center appear likely at this writing.

The Ultimate Objectives of SOBE-R.

Earlier we stressed that SOBE-R has been under evaluation since its inception. That is, there has been a continuous process of try-out and feedback as is reflected in part in our formalized field-testing agreements. This is the type of evaluation Scriven refers to as "formative." Soon, of course, we will be in a position to begin to deal with more ultimate kinds of evaluative questions. In other words, given the fact that the SOBE-R system has been carefully developed and is ready for implementation on a large scale, what kinds of differences should it make? What "summative" effects

[3]Krathwohl, D.R. Stating Objectives Appropriately for Program, for Curriculum, and for Instructional Materials Development. *Journal of Teacher Education*, March, 1965, 83-92.

will SOBE-R have if it is successful?

. Just how one might go about evaluating an evaluation system is an interesting question. Standardized tests have been "evaluated" in terms of technical considerations such as reliability, content validity, etc. These issues certainly pertain to tests derived from SOBE-R item pools (for example, the congruency study discussed earlier) but do not in any way get at the potential impact of SOBE-R. The summative effects of educational programs are typically evaluated in experiments or quasi-experiments with comparisons between groups of learners or, where such comparison groups cannot be found, by means of designs involving statistical controls. But SOBE-R is not an "instructional" program. Its potential effects on the performance of learners lie at the end of a long chain of effects on others, such as planners, curriculum developers, administrators, teachers, and even parental groups.

If we cannot evaluate SOBE-R in the way that contemporary tests are evaluated or in the ways that instructional programs are evaluated, what can we do? Answers to this question must devolve from the particular kinds of effects SOBE-R is expected to have. We see SOBE-R influencing education over the long term in three ways.

(1) *The availability of, and committment to, an explicit set of objectives should generate clearer and more precise understanding of the goals of instruction at all levels of the education profession and in the mind of the public and its representatives.* Such understanding of the goals of instruction is important for several reasons. First, without detailed specification of objectives it is impossible to set meaningful standards for what is to be accomplished. There can be no context of accountability if it is impossible to specify definitively what it is that the schools are accountable for. Second, it is vital to be aware of the desired learning outcomes at a given level when designing curricula, selecting materials, etc. This major planning use of SOBE-R has already been mentioned. Thirdly, the possibility of meaningful and informative communication between schools and the public that supports the schools depends upon a common understanding of what the schools are committed to accomplishing. Today we have virtually nothing but percentile ranks on standardized tests and the grading system, which are widely misunderstood both by the public and members of the educational profession. Tomorrow we hope to have a publicly available set of desired learning outcomes. This is evident, for example, in the plans of the Florida State Department of Education already mentioned.

Given the above, it is clear that one important type of evaluative information pertaining to SOBE-R will consist of evidence of the extent to which there is knowledgeability among members of the education profession and the concerned public about the selected instructional objectives. Such evidence is complex, but would be manifest in many ways. For example, are teachers aware of the objectives at which an instructional unit is directed? Were the objectives considered in the selection of materials such as textbooks, reading exercises, and the like? Is the learning performance of children reported to parents in terms of objectives accomplished during the school year? Are references to the objectives a salient part of the dialogue among teachers, principals, and others? These and many other kinds of indices would all indicate the extent to which the availability of the SOBE-R objectives has shaped resource allocation, planning, and communication.

(2) *Educational agencies using SOBE-R will utilize assessment information in determining needs, planning reading programs, and managing the instructional process.* Determining the extent to which assessment information is utilized also poses a complex problem in data collection.. But unless it can be demonstrated that the information generated actually enters into decision making about programs of reading instruction, it cannot be maintained that SOBE-R is successful. It is clear that the information from assessments of learner progress is pertinent at many levels of decision making. In Florida, for example, will funding requests to legislative committees use assessment results as justification? Will this information be deemed pertinent by legislators and other resource allocators? Will priorities be established at all levels partly on the basis of assessment results? Will schools utilize assessment results in planning instruction and communicating to the community? Procedures used to collect such information would obviously be various, including interviews, inspection of documents, and questionnaires.

(3) *Ultimately, the reading achievement of children in educational agencies utilizing SOBE-R will show incremental impprovement.* Here we have the real purpose of SOBE-R. This criterion of student achievement can be met only if (1) and (2) above are met. It is also dependent upon the wisdom of the decisions that are made and the effort directed at carrying out those decisions. In this sense, any impact of SOBE-R on the performance of children is filtered through and moderated by a host of independent causal factors. The unique influences of SOBE-R as a planning and management tool on student learning cannot be isolated, but it seems to us that the question must be asked anyway.

It does not appear likely that the matter of student achievement can be dealt with through evaluative research comparing agencies using SOBE-R with agencies not using the system. Rather, time-series designs tracing levels of reading achievement over time within agencies utilizing SOBE-R would be likely to yield the most pertinent information. Research utilizing the former or comparison group approach would be obfuscated by initial differences between users and non-users and would not be able to maintain control over what non-users decide to do during the period of the study. This is a problem which plagues most evaluative studies attempting to apply experimental paradigms to educational phenomena.

The evaluation of a system like SOBE-R has been presented in a way that is consistent with the evaluation model advanced by the Center for the Study of Evaluation. We have attempted to show that evaluation is a continuous process, beginning with an assessment of need and proceeding through formative stages, each requiring different kinds of input. The development, tryout, and revision of SOBE-R materials will receive most of our attention for the remainder of this year and much of the next. But in Fall 1971 initial administrations of SOBE-R assessment instruments will be conducted by our field-testing agencies.

The resulting availability of information on student achievement means that summative effects of the types described in the last section may begin to occur. We need to be in a position to note and record these effects, even though no claim can be made that a final summative evaluation of SOBE-R can or should occur during 1971-1972. Rather, we would be primarily concerned with identifying indices that reflect such effects for later use in a summative evaluation of the operational system.

FILMED FROM BEST AVAILABLE COPY